

A Web GIS-Based Platform to Harvest Georeferenced Data from Social Networks: Examples of Data Collection Regarding Disaster Events

<https://doi.org/10.3991/ijoe.v14i02.8191>

Cidália Costa Fonte, Diogo Fontes, Alberto Cardoso^(✉)
University of Coimbra, Portugal
alberto@dei.uc.pt

Abstract—Whenever disaster situations occur the civil protection authorities need to have fast access to data that may help to plan emergency response. To contribute to the collection and integration of all available data a platform that aims to harvest Volunteered Geographical Information (VGI) from social networks and collaborative projects was created. This enables the integration of VGI with data coming from other sources, such as data collected by physical sensors in real time and made available through Applications Programming Interface (APIs), as well as, for example, official maps. The architecture of the created platform is described and its first prototype presented. Some example queries are performed and the results are analyzed.

Keywords—Volunteered Geographic Information (VGI), physical sensors, disaster management, Geographic Information Systems (GIS)

1 Introduction

The emergency response in case of extreme events requires the collection of data about the event itself before the authorities arrive at the site. This includes not only information about its location and the most appropriate route to reach it, but also data about, for example, its magnitude or the presence of populations injured or in danger. Information about the region surrounding the location of the event may also be valuable to assess its seriousness or the risk of occurrence of secondary hazards. These may be, for example, the location of industries with highly inflammable products in the region where a fire deflagrated, the existence of populations in the neighborhood with mobility restrictions or infrastructures hosting a large number of citizens.

The data about the event itself may be difficult to collect in real time by the authorities, especially at its initial stage or if it has a large spatial scale. Therefore, the data collected by citizens that are in the vicinity of the occurrence and are shared in real time in social networks, such as Flickr, Twitter, Facebook or Instagram, may be valuable to collect as much data as possible to prepare an appropriate response. Moreover, some of the data required about the region may not be available in official maps, such as, for example, the location of restaurants, bars or factories. However, some of these

data are available in collaborative projects such as OpenStreetMap (OSM), which are continuously being updated by the citizens.

The identification of all relevant crowdsourced georeferenced data, termed Volunteered Geographic Information (VGI) [1], requires searching in each project and then integrating all the data extracted from different projects, which may have different characteristics and formats. However, this will take time and a key factor for an efficient emergency response is celerity. Therefore, the main motivation for the creation of the web GIS-based platform described in this article was to create a tool enabling easy harvesting of VGI from social networks and collaborative projects, and its immediate integration in the same platform, along with other types of georeferenced data, such as data collected by physical sensors and official maps. This will enable the fast and easy identification of data useful to assist authorities, without requiring any knowledge about the collaborative projects from which the data was obtained.

The difference of this platform when compared to similar ones created within other projects is that it may collect data from any VGI project that provides an Application Programming Interface (API) enabling a search by location, instead of requiring a dedicated app for data collection, as in [2]. This was also suggested by [3], but for a different aim, which was the collection of Land Use/Land Cover data. The integration of all this data into the same system will also allow validation procedures to be implemented, in order to filter data with the highest chance of providing useful information. This is particularly relevant for emergency, where the time spent filtering relevant information among all data available should be minimized, as time is a crucial variable for emergency response.

2 Platform design

The created platform aims to collect and integrate data from several sources. The architecture of the proposed platform is structured considering three tiers: a data tier, an intermediate tier and an application tier (see Figure 1). The data tier includes all data sources considered, which may be social networks, other VGI projects, data provided by physical sensors or any other existing geospatial data. The intermediate tier includes readers for each data source, which may provide different types of data with different characteristics and therefore need to be translated for data extraction from the projects. This tier also includes an integration and processing component, a database and adapters that will make the data and services available for each type of user. The application tier includes all types of users that will have access to different tools and functionalities.

The platform is designed to work with two different kinds of loops. One is designed to run continuously, searching for predefined types of events, such as fires or floods, and will enable the identification of bursting events. This requires the previous identification of a geospatial location of interest and a set of relevant keywords for each type of event. The obtained data may be discarded if considered not relevant or stored in the database for latter access and analysis. The second loop aims to perform manual searches, specifying a search location and any keywords of interest. The out-

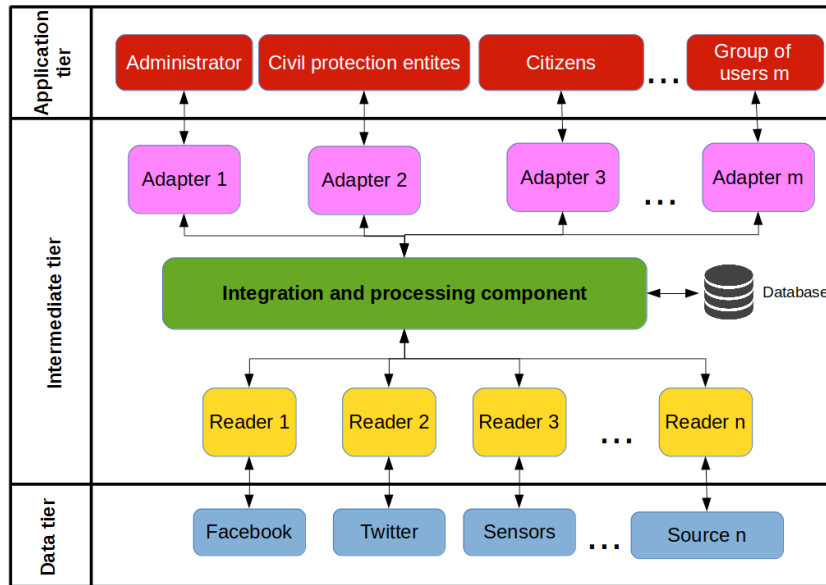


Fig. 1. Platform architecture diagram [4].

outputs are visualized automatically as point features over a map or imagery and may be stored in the database. Clicking in the point will enable the visualization of the data stored in the database and any associated photographs or videos. This mode enables the user to perform any type of search of interest to collect additional data about an event already identified.

As the extraction of data from some VGI projects, such as OSM, may take some time, especially if a large area is at stake, the continuous loop will enable the extraction and storage of data considered potentially relevant, so that it can be latter retrieved from the database without requiring a search in real time, which might increase the response time.

3 Platform use and search results

The use of the platform will be illustrated by searching for a set of specific keywords and location to harvest data from social networks. For this aim the platform requires:

1. The identification of a region to perform the search. This choice is made placing a marker over the map visible in the application and selecting a search radius that is centered on the marker;
2. The selection of the social networks and VGI projects where the search will be performed. In the prototype shown in Figure 2 it is possible to search in OSM, Flickr and/or Twitter.

3. The selection of a set of keywords to perform the search in the social networks. If no keywords are selected all data will be retrieved, until the maximum number of data that can be downloaded in each source application is reached. For OSM a *Feature* type needs to be identified from a drop-down list. A *Value* may also be chosen for the selected *Feature*, but it's choice is optional. If no *Value* is selected, all elements of the selected *Feature* will be extracted.
4. The number of posts that will be extracted from each social network (optional).

Figure 2 shows the example of a search operation centered at Grenfell Tower (London), with a radius of 1 km, where elements corresponding to the *Feature* = *building* and *Value* = *apartments* are requested from OSM, and the keyword *fire* is used to search for posts in Flickr.

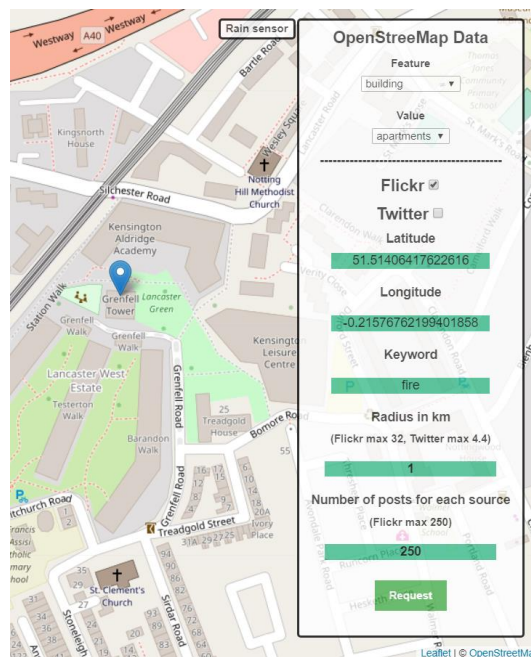


Fig. 2. Example of a search operation: 1) placement of a marker over Grenfell Tower, 2) selection of: a) a search radius of 1 km, b) apartment buildings in OSM and c) posts related to *fire* in Flickr.

Once the platform finishes the search the results may be visualized over OSM or a satellite image. The identified posts are aggregated in clusters (represented by a circle with a central number indicating the number of contained posts). The clusters are created or disaggregated according to the map scale used in the visualization.

Figure 3 shows the results obtained for the search represented in Figure 2. The polygons extracted from OSM are represented with the blue outline (three blocks in the visible area). A few dozens of photographs were obtained in the region of the Grenfell Tower. The white labels with numbers identify the location of the photographs shown



Fig. 3. Results of the search illustrated in Figure 2. Polygons with the blue outline were extracted from OSM (apartment buildings). White labels were placed over the image to enable an identification of the photographs shown in Figure 4.

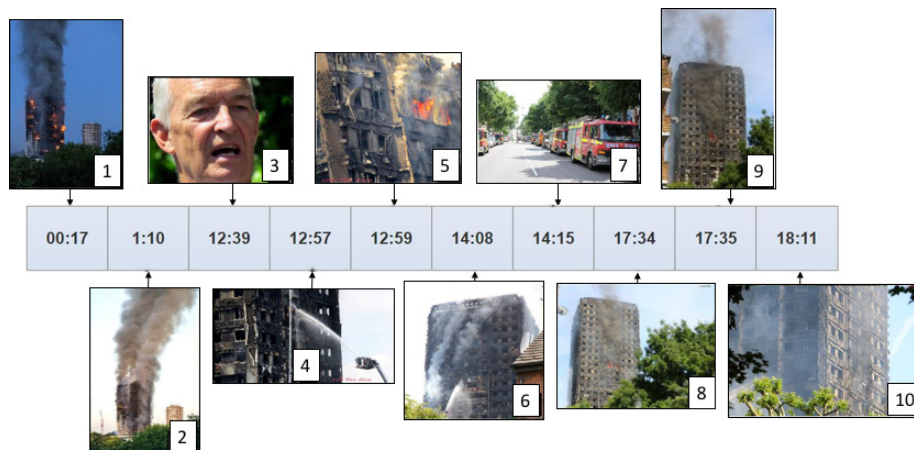


Fig. 4. Retrieved photographs from Flickr posts labeled in Figure 3, along with a timeline of photographs collection.

in Figure 4, where the field *datetaken* associated to the photographs was used to create a timeline linked to the downloaded photographs.

The results show that photographs 1 and 2 were posted with a *datetaken* time of respectively 00:17 and 01:10. It can be easily seen that these photographs were not taken during the night but already in the morning, so the clock of the camera used to take the photographs (which were taken by the same volunteer) did not have the correct time set. On the other hand, some photographs are not useful for the aim of the application. For example, photographs labeled as 3 and 7 do not include information useful for the authorities. Therefore, to optimize the time spent searching for useful data when lots of data will be available, a filtering process will be implemented in the platform. Details about the filtering system will be provided in future publications.

Figures 5, 6 and 7 illustrate another example of data collected for Calgary, regarding the severe floods that occurred in 2013. The search shown in OSM aims to retrieve commercial areas (*Feature = shop*) and the selection of photographs from Flickr using the keyword *flood*. A search radius of 1km was used around the point placed over Bow River (Figure 5).

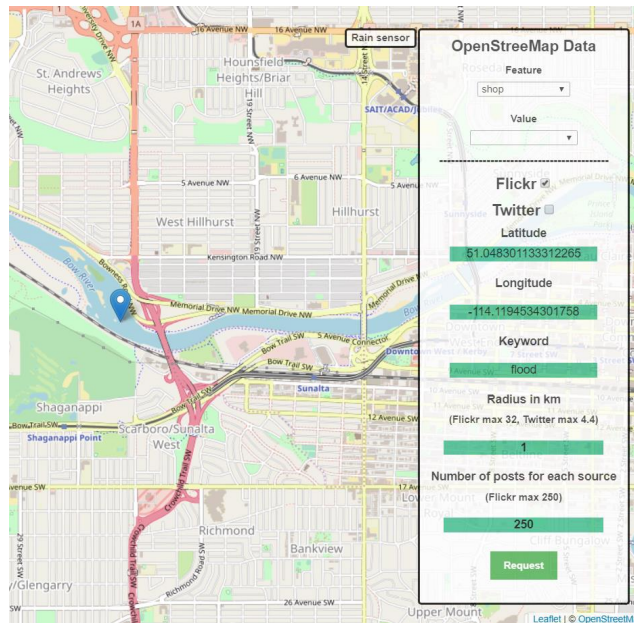


Fig. 5. Example of a search operation: 1) placement of a marker over Bow River, 2) selection of: a) a search radius of 1km, b) shops in OSM and c) posts related to *flood* in Flickr.

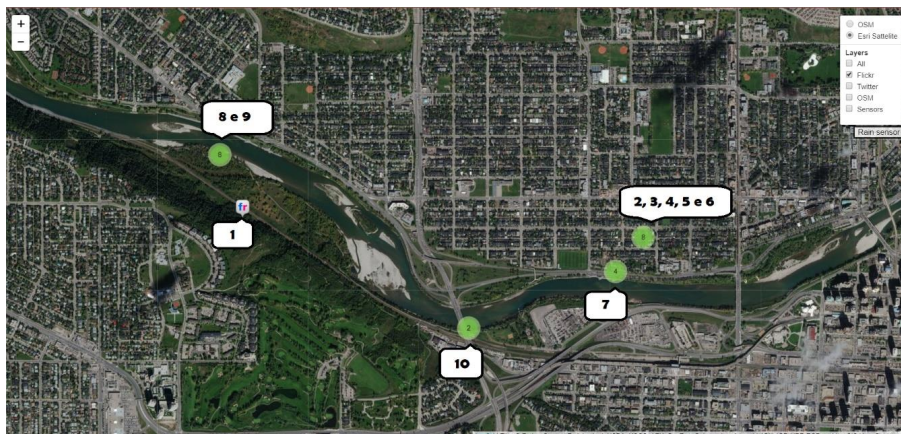


Fig. 6. Results of the search illustrated in Figure 5 for the Flickr photographs. White labels were placed over the image to enable an identification of the photographs shown in Figure 7.

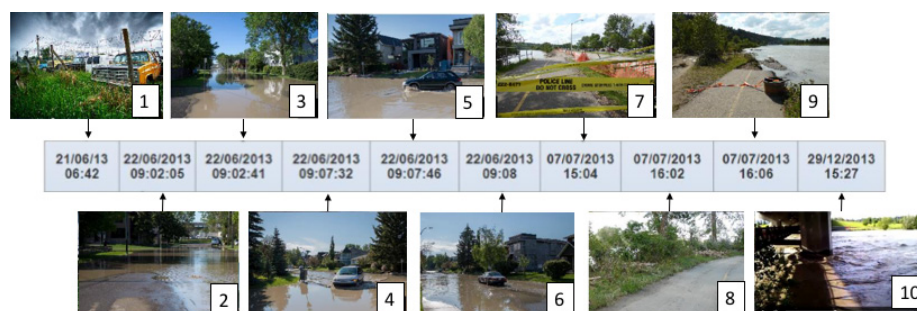


Fig. 7. Retrieved photographs from Flickr posts labeled in Figure 6, along with a timeline of photographs collection.

Figure 6 and 7 show the results obtained for the photographs extracted from Flickr. The photographs have dates between the 21st of June and the 7th of July 2013, and the last image shows a video from the 29th of December 2013. It can also be seen that the photograph labeled as 1 does not show any floods, and is therefore useless for flood analysis. The other photographs and the video shown in 10 either show cases of flood or damages caused by the floods that occurred in the area in 2013.

4 Conclusions

The prototype of the platform presented in this article has the possibility to extract georeferenced data from social networks and projects that provide an API for the extraction of VGI, insert the information into a database, and show the results over a map or image. Other sources of data are available for extraction in the platform, such as data from meteorological stations, as well as official maps. The results showed that relevant data may be obtained from the platform, even though several improvements need to be implemented, which include: 1) enable the search in the social networks by date and hour of data collection; 2) add other data sources of VGI, such as, for example, posts from Facebook and Instagram, to increase the amount of data available for analysis; 3) implement a validation and filtering process to avoid the visualization of useless and erroneous data.

In future versions of the platform, different user profiles will be implemented, providing access to different datasets and functionalities, allowing the use of the platform for a wider variety of situations, as the implemented functionalities may be useful for other types of applications.

5 Acknowledgments

The authors acknowledge the support of INESC Coimbra under project INESCC FR004 (ExtremeCGI - Monitoring Extreme Events integrating Crowdsourced Geographic Information and Real-time Sensor Data) and Fundação para a Ciência e a Tecnologia (FCT) under project grant UID/MULTI/00308/2013.

6 References

- [1] Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. *Geo-Journal*, 69 (4): 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- [2] Horita, F., de Albuquerque, J.P., Degrossi, L., Mendiando, E. and Ueyama, J. (2015). Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks. *Computers & Geosciences* 80: 84-94. <https://doi.org/10.1016/j.cageo.2015.04.001>
- [3] Estima, J., and Painho, M. (2016). User generated spatial content-integrator: conceptual model to integrate data from diverse sources of user generated spatial content. *ISPRS Int. J. Geo-Inf.*, 5(10): 183. <https://doi.org/10.3390/ijgi5100183>
- [4] Fontes, D., Fonte, C.C., Cardoso, A. (2017). "Integration of VGI and sensor data in a Web GIS-based platform to support emergency response". 4th Experiment@ International Conference, Faro, Portugal, Publisher: IEEE, pp. 214-219. <https://doi.org/10.1109/EXPAT.2017.7984345>

7 Authors

Cidália Costa Fonte has a PhD in Geographical Engineering and is an Assistant Professor at the Mathematics Department of the Faculty of Sciences and Technology, University of Coimbra, Portugal. She is also a researcher and member of the board of Directors of the Institute of Systems Engineering and Computers at Coimbra and the coordinator of the Master course in Geospatial Information Engineering, Faculty of Sciences and Technology, University of Coimbra, Portugal. (email: cfonte@mat.uc.pt)

Diogo Fontes has a bachelor degree in Environmental Engineering from Technical University of Lisbon and is MSc in Geographic Information Technologies by University of Coimbra, Portugal. (email: dds.fontes@gmail.com)

Alberto Cardoso received the Ph.D. degree in Informatics Engineering from the University of Coimbra, is Assistant Professor with Tenure at the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra and senior member of the Centre for Informatics and Systems of the University of Coimbra (CISUC) and of the International Society for Engineering Pedagogy (IGIP). He was awarded the title of "International Engineering Educator Honoris Causa" ("Ing.Paed.IGIP h.c.") by IGIP in 2016. (email: alberto@dei.uc.pt)

Article submitted 03 January 2018. Final acceptance 06 February 2018. Final version published as submitted by the authors.